

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/160539>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

ON A KINETIC OPINION FORMATION MODEL FOR PRE-ELECTION POLLING

 **Bertram Düring**

Mathematics Institute
University of Warwick
Zeeman Building
Coventry, CV4 7AL
United Kingdom

bertram.during@warwick.ac.uk

 **Oliver Wright**

Mathematics Institute
University of Warwick
Zeeman Building
Coventry, CV4 7AL
United Kingdom

oliver.l.j.wright@warwick.ac.uk

ABSTRACT

Motivated by recent successes in model-based pre-election polling, we propose a kinetic model for opinion formation which includes voter demographics and socio-economic factors like age, sex, ethnicity, education level, income and other measurable factors like behaviour in previous elections or referenda as a key driver in the opinion formation dynamics. The model is based on *Toscani's* kinetic opinion formation model [1] and the leader-follower model of *Düring et al.* [2], and leads to a system of coupled Boltzmann-type equations and associated, approximate Fokker-Planck-type systems. Numerical examples using data from general elections in the United Kingdom show the effect different demographics have on the opinion formation process and the outcome of elections.

Keywords opinion formation, demographics, pre-election polling, Boltzmann equation, Fokker–Planck equation.

1 Introduction

Different kinetic models to describe opinion formation in a society consisting of a large number of interacting individuals have been proposed and analysed in the literature, see e.g. [1–25]. Building on mathematical approaches from statistical mechanics they frequently lead to generalizations of the classical Boltzmann equation for gas dynamics. Instead of binary collisions, one considers the process of compromise between two individuals. Similar models are used in the modelling of wealth and income distributions which show Pareto tails, cf. [26] and the references therein. A key application of such kinetic opinion formation models is political opinion formation in elections or referenda. When aiming to push the boundaries of where such methods can be applied to analyse or forecast election results (see also [27]) it is useful to look at recent developments in pre-election polling.

Pre-election polling which aims to forecast election outcomes based on surveys of political opinion used to rely on relatively small samples. But a small number of observations at national level make it difficult to forecast outcomes at sub-national (constituency) level, leading to frequent (real or perceived) failures of pre-election polls [28], e.g. in 2015 and 2017 UK general elections. Even if a pre-election poll correctly forecasts the share of a party at national level, who wins the election in a first-past-the-post electoral system depends on the outcome on the local (constituency) level, i.e. who wins each of the many seats in parliament. In recent years, so-called *multilevel regression and post-stratification* (MRP) methods have shown to be able to form high-quality sub-national estimates of political opinions by uncovering dependencies of voter behaviour on demographic, socio-economic and other measurable factors [29]. MRP methods are much better at capturing the complex nuances of opinions depending on local variations in historic voter behaviour and voter demographics. To this end they use a larger poll size so that sufficient samples from each constituency are gathered, and at the same time record demographic and socio-economic characteristics like age, sex, ethnicity, education level, income and information on previous votes in elections or referenda from the participants. Then a regression is performed to determine how likely it is that an individual with a certain combination of demographic and socio-economic characteristics votes for a particular party. In a post-stratification step, census and other publicly available data is used to determine the composition of the voter population in each constituency with

regard to these demographic and socio-economic characteristics and then forecast the outcome of the vote in each of the constituencies.

In this paper, motivated by the development and successes of MRP methods for pre-election polling, we propose and analyse a kinetic model for opinion formation which takes into account demographic and socio-economic factors like age, sex, ethnicity, education level, income and other measurable factors like behaviour in previous elections or referenda. The basic underlying kinetic model opinion formation is the leader-follower model from [2] which in turn is based on the opinion formation model by Toscani [1]. Mathematically, it is related to works in the kinetic theory of granular gases [30]. In the kinetic opinion formation models the evolution of the distribution of opinion is described by means of microscopic interactions among individuals in a society. These microscopic interactions are reminiscent of collisions in gas dynamics. Opinion is represented as a continuous variable $w \in \mathcal{I}$ with $\mathcal{I} = [-1, 1]$ representing the left–right political spectrum, and ± 1 represent extreme opinions. In each interaction two main drivers affect an individual’s opinion. The first one is a compromise process [31–33], in which individuals tend to reach a compromise after exchange of opinions. The second one is self-thinking, where individuals change their opinion in a diffusive way, possibly influenced by exogenous information sources such as the media. In this setting, the time evolution of the distribution of opinion among individuals is governed by (systems of) Boltzmann-type equations. In a suitable quasi-invariant limit, partial differential equations of Fokker–Planck type can be derived for the distribution of opinion.

The paper is organized as follows. In the next section we propose a generalised kinetic opinion formation model which takes into account demographic and socio-economic factors in the compromise process of the kinetic opinion formation model, and present the resulting system of Boltzmann-type equations and its associated quasi-invariant limit Fokker-Planck system. In Section 2 we propose an approach to use real-world pre-election polling and census data to inform the choice of compromise functions in our model. We put this to test in two real-world examples from the 2019 United Kingdom general elections, for the East of England region and Nottinghamshire which forms part of the so-called “red wall” of constituencies in northern England which was widely perceived to have played a key role in the outcome of the 2019 elections. We present results from direct Monte Carlo simulations of the microscopic model and compare it with finite element based simulations of the Fokker-Planck system. Section 4 concludes.

2 Derivation of the Model

2.1 Interactions

We propose a model based on *Toscani’s* opinion formation model [1] and the leader-follower model of *Düring et al.* [2], where we look at the distribution of the individuals’ opinion, w , in an interval $\mathcal{I} = [-1, 1]$. We split our population into N follower species and a leader species, which we label as species $1, 2, \dots, N$ and species L and associate with a density function, $f_i(w, t)$ and $f_L(w, t)$, respectively, depending on the opinions, $w \in \mathcal{I}$, and time, $t \geq 0$. Each follower species will be associated with a single demographic. Each demographic can include numerous demographic and socio-economic factors like age, sex, ethnicity, education level, income and other measurable factors like behaviour in previous elections or referenda. To limit the number of species to a practical value, we do not include all possible demographic factors, but consider only the factors that are assumed to be most important in driving opinion formation [34].

The dynamics of our system are defined by looking at binary interactions (or collisions) between two individuals from our total population. We define the interactions in this model, in three possible cases: (a) the interactions where a member of the leader species interacts with another member of the leader species (an L and L interaction); (b) a member of the leader species interacts with a member of the follower species i for some $i = 1, \dots, N$ (an L and i interaction); and (c) a member of a follower species interacts with another member of a follower species (an i and j interaction for some $i, j = 1, \dots, N$). We assume that the leader species is formed by strong opinion leaders, as in [2], that is assertive individuals able to influence followers but withstand being influenced themselves. To that end, we have that a leader species will have their opinion unchanged by an interaction with an individual from a follower species, but interact normally with another leader. Thus we look at the following cases for the interactions with w_L^*, v_L^* and w_L, v_L as the post- and pre-interaction leader opinions, respectively w^*, v^* and w, v as the post- and pre-interaction opinions of the follower species, respectively:

(a) The L and L case:

$$\begin{aligned} w_L^* &= w_L + \gamma_L P_{LL}(w_L, v_L)(v_L - w_L) + D_L(w_L)\eta_L, \\ v_L^* &= v_L + \gamma_L P_{LL}(v_L, w)(w_L - v_L) + D_L(v_L)\eta_L; \end{aligned} \tag{1}$$

(b) The i and L case, $i = 1, 2, \dots, N$:

$$\begin{aligned} w^* &= w + \gamma_i P_{iL}(w, v_L)(v_L - w) + D(w)\eta_i, \\ v_L^* &= v_L; \end{aligned} \quad (2)$$

(c) The i and j case, $i, j = 1, 2, \dots, N$:

$$\begin{aligned} w^* &= w + \gamma_i P_{ij}(w, v)(v - w) + D(w)\eta_i, \\ v^* &= v + \gamma_j P_{ji}(v, w)(w - v) + D(v)\eta_j. \end{aligned} \quad (3)$$

In the interactions above, $\gamma_\nu \in \mathbb{R}$, for $\nu = 1, 2, \dots, N, L$, denotes compromise parameters which govern the “speed” of attraction of two different opinions, the functions $P_{ij}(\cdot, \cdot)$ and $D(\cdot)$ model the local relevance of compromise and self-thinking for a given opinion, respectively, and η_ν are the “self-thinking” random variables, which model how the opinion of an individual changes independently from other individuals. Here, we have made the implicit assumption that the total number of followers and leaders in a species – and by extension the mass of the density functions – remain constant over time. We only mention here that when using this to model a phenomenon with a timescale of years, demographic changes (migration to other species, birth/death) could also be incorporated by adding gain and loss terms similarly to [2].

To ensure that w^* and v^* remain in \mathcal{I} , we choose P_{ij} and D functions accordingly. Our P_{ij} functions are chosen in such a way that they remain in $[0, 1]$. In the prevalent literature, this has been taken to be a localisation function, such as a characteristic function or a hyperbolic tangent, [1, 2, 6], here we propose a more realistic function based on real-world polling data in Section 2.2.3 below. Then, $P_{ij}(w, v)(v - w) \in [-2, 2]$, hence we choose $\gamma_\nu \in (0, 1/2)$, so that $-1 < \gamma_\nu P_{ij}(w, v)(v - w) < 1$. The function D is chosen in such a way as to ensure that the diffusion term is sufficiently small near the boundaries which guarantees that the post-interaction opinions remain in \mathcal{I} . Thus we require the function D to remain in $[0, 1]$ and $D(-1) = D'(-1) = D(1) = D'(1) = 0$. We also want D to be symmetric in this case, although this is not required. A usual choice for this function is:

$$D(x) = (1 - x^2)^\alpha,$$

for some $\alpha > 0$. We choose D_L similarly and for both $\alpha = 2$, unless otherwise stated.

Finally, we now discuss the random variables η_ν . Let us define a set of probability measures in the following way, for some $\delta > 0$:

$$\mathcal{M}_{2+\delta} = \left\{ \mathbb{P} : \mathbb{P} \text{ is a probability measure and, } \langle |\eta|^\alpha \rangle = \int_{\mathcal{I}} |\eta|^\alpha d\mathbb{P}(\eta) < \infty, \forall \alpha \leq 2 + \delta \right\}.$$

We choose a probability density $\Theta \in \mathcal{M}_{2+\delta}$ and we assume that this density is obtained from a random variable Y with the properties, $\langle Y \rangle = \int_{\mathcal{I}} Y d\Theta(Y) = 0$, $\langle Y^2 \rangle = \int_{\mathcal{I}} Y^2 d\Theta(Y) = 1$, then we choose η_ν with the property that $\langle |\eta_\nu|^p \rangle = \langle |Y \sigma_\nu|^p \rangle$ for all $\nu = 1, 2, \dots, N, L$ and $p \in [0, 2 + \delta]$, with $0 < \sigma_\nu < 1$. We assume that the opinion of an individual is more likely to be swayed by an interaction with another individual rather than their exposure to outside factors, thus $\sigma_\nu \ll 1$.

2.2 Boltzmann-type and Fokker-Planck equations

From the microscopic interactions in the previous subsection $N + 1$ Boltzmann-type equations can be derived using standard methods of kinetic theory [2, 35]. These come in two forms, N equations that describe the behaviour of the follower species,

$$\frac{\partial}{\partial t} f_i(w, t) = \frac{1}{2\tau_{iL}} Q(f_i, f_L)(w, t) + \sum_{j=1}^N \frac{1}{2\tau_{ij}} Q(f_i, f_j)(w, t), \quad \text{for } i = 1, 2, \dots, N, \quad (4)$$

and the equation that describes the behaviour of leader species,

$$\frac{\partial}{\partial t} f_L(w, t) = \frac{1}{2\tau_L} Q(f_L, f_L)(w, t), \quad (5)$$

with the interaction operators $Q(f_\mu, f_\nu)(w, t)$ between species $\mu, \nu = 1, \dots, N, L$ in (4) and (5), expressed in weak form as

$$\int_{\mathcal{I}} Q(f_\mu, f_\nu)(w, t) \phi(w) dw = \left\langle \int_{\mathcal{I}^2} [\phi(w^*) + \phi(v^*) - \phi(w) - \phi(v)] f_\mu(w, t) f_\nu(v, t) dw dv \right\rangle,$$

where ϕ are test functions from $C_c(\mathcal{I}; \mathbb{R})$, that are continuous and compactly supported functions on \mathcal{I} . In (4) and (5), we use τ_{ij} , τ_{iL} and τ_L to denote the relaxation times of the Boltzmann-type equations which allow to control the interaction frequencies.

We use similar methods to those described in [2] to derive a set of Fokker-Planck equations for this model by taking a quasi-invariant limit of (4), (5). It is well-known that these quasi-invariant limits are a good approximation for the stationary profiles of the Boltzmann-type equation [1]. To study these large time stationary distributions, therefore, we use the transformation $s = \gamma t$, for some $\gamma \ll 1$, writing the large time transformed density functions $g_\mu(w, s) = f_\mu(w, t)$ for every μ , using the scaling $\gamma_\nu = \alpha_\nu \gamma$ and $\sigma_\nu = \beta_\nu \sigma$ for all $\nu = 1, \dots, N, L$, and restricting our test functions to twice differentiable functions, all of whose derivatives are δ -Hölder continuous for some $\delta > 0$ with compact support, that is $\phi \in C_c^{2,\delta}(\mathcal{I}, \mathbb{R})$. In the limit $\gamma, \sigma \rightarrow 0$ while keeping $\lambda = \sigma^2/\gamma$ fixed, this yields the following nonlinear, nonlocal Fokker-Planck equations for the follower species,

$$\begin{aligned} \frac{\partial}{\partial s}(g_i(w, s)) = & \frac{\partial}{\partial w} \left[\left(\frac{\alpha_i}{2\tau_{iL}} \mathcal{K}_{L,iL}(w, s) + \sum_{j=1}^N \frac{\alpha_j}{2\tau_{ij}} \mathcal{K}_{j,ij}(w, s) \right) g_i(w, s) \right] \\ & + \sum_{j=1}^N \frac{\alpha_j}{2\tau_{ij}} \frac{\partial}{\partial w} \left(\mathcal{K}_{i,ij}(w, s) g_j(w, s) \right) \\ & + \left(\frac{\lambda_i M_L(s)}{4\tau_{iL}} + \sum_{j=1}^N \frac{\lambda_j M_j(s)}{4\tau_{ij}} \right) \frac{\partial^2}{\partial w^2} \left(D^2(w) g_i(w, s) \right) \\ & + \sum_{j=1}^N \frac{\lambda_j M_i(s)}{4\tau_{ij}} \frac{\partial^2}{\partial w^2} \left(D^2(w) g_j(w, s) \right), \quad \text{for } i = 1, 2, \dots, N, \end{aligned} \quad (6)$$

and, similarly, the Fokker-Planck equation for the leader species L :

$$\begin{aligned} \frac{\partial}{\partial s}(g_L(w, s)) = & \frac{\alpha_L}{\tau_L} \frac{\partial}{\partial w} \left(\mathcal{K}_{L,LL}(w, s) g_L(w, s) \right) \\ & + \frac{\lambda_L M_L(s)}{2\tau_L} \frac{\partial^2}{\partial w^2} \left(D_L^2(w) g_L(w, s) \right), \end{aligned} \quad (7)$$

where $\lambda_\nu = \beta_\nu^2 \lambda$ for all $\nu = 1, \dots, N, L$ and with $\xi, \mu, \nu = 1, 2, \dots, N, L$,

$$\mathcal{K}_{\xi,\mu\nu}(w, t) := \int_{\mathcal{I}} P_{\mu\nu}(w, v)(v - w) g_\xi(v, t) \, dv, \quad M_\xi(s) := \int_{\mathcal{I}} g_\xi(w, s) \, dw.$$

We now show that the density masses remain constant for all $s > 0$ in the Fokker-Planck model, that is $M_\zeta(s) = M_\zeta(0)$ for all $s > 0$ and $\zeta = 1, 2, \dots, N$. Consider the following by substituting the weak form of (6) into $\frac{d}{ds} \int_{\mathcal{I}} g_\zeta(w, s) \phi(w) \, dw$,

$$\frac{d}{ds} \left(M_\zeta(s) \right) = \frac{d}{ds} \int_{\mathcal{I}} g_\zeta(w, s) \phi(w) \, dw \quad (8)$$

$$\begin{aligned} = & \int_{\mathcal{I}} \left(\frac{1}{2\tau_{\zeta L}} \mathcal{K}_{L,L\zeta}(w, s) + \sum_{j=1}^N \frac{1}{2\tau_{\zeta j}} \mathcal{K}_{j,\zeta j}(w, s) \right) g_\zeta(w, s) \phi'(w) \, dw \\ & + \int_{\mathcal{I}} \sum_{j=1}^N \frac{\alpha_{\zeta j}}{2\tau_{\zeta j}} \mathcal{K}_{\zeta,\zeta j}(w, s) g_j(w, s) \phi'(w) \, dw \\ & + \int_{\mathcal{I}} \left(\frac{\lambda_\zeta M_L(s)}{4\tau_{\zeta L}} + \sum_{j=1}^N \frac{\lambda_\zeta M_j(s)}{4\tau_{\zeta j}} \right) D^2(w) g_\zeta(w, s) \phi''(w) \, dw \\ & + \int_{\mathcal{I}} \sum_{j=1}^N \frac{\beta_{\zeta j}^2 \lambda_\zeta M_\zeta(s)}{4\tau_{\zeta j}} D^2(w) g_j(w, s) \phi''(w) \, dw. \end{aligned} \quad (9)$$

A choice for the test function, consistent with the conditions we have placed on it, is $\phi(w) \equiv 1$, which leads directly to $\frac{d}{ds}(M_\zeta(s)) = \frac{d}{ds} \int_{\mathcal{I}} g_\zeta(w, s) \, dw = 0$ and thus to $M_\zeta(s) = M_\zeta(0)$ for all $s > 0$. Analogously, we obtain $M_L(s) = M_L(0)$ for all $s > 0$.

From this point, it will be prudent to rename the function $g_\nu(w, s)$ back to $f_\nu(w, t)$ for all $\nu = 1, 2, \dots, N, L$ for ease of presentation.

2.3 Choice of compromise function from demographic voter intention data

In [1, 2, 6], the compromise function P has been chosen as a mere localisation function, usually a hyperbolic tangent function or a characteristic function for some radius r . Although this approach has been used to great effect when considering two party systems [6], we propose here a more data-driven approach to include demographic and other factors into the compromise functions, which is more suited to a multi-party political system. To achieve this, we look to voter intention polls made before the election, thus, as briefly explained in Section 2.2.1, we propose a new form of P which is based on that polling data. Each follower species will be associated with one combination of n “characteristics”, $c_{i1}, c_{i2}, \dots, c_{in}$, each of which is chosen from a potential set of mutually exclusive groups, that is all individuals in a species belong to one, and only one, group, and we assign to all our species one of these groups. Using the UK General Elections as an example, we could choose the characteristics of age and home-ownership. These have mutually exclusive groups that could be associated with them, that is under 24 years-old, and over 24 years old for age, and either a home-owner or not for home-ownership. We then use four species to which we assign the groups, under 24 home-owners, over 24 home-owners, under 24 non-home-owners and over 24 non-home-owners. To this end, we make the assumption that $P_{ij}(w, v)$ can be decomposed into a localisation function P_{loc} , which depends only on the opinions, and characteristic comparison functions P_{char} , which depend on the opinions and, in addition the data from pre-election polls [36] for each of the characteristics of the species i and j .

Let w, v be the opinions of two individuals in species i and j , respectively, with associated characteristic groups c_{ik}, c_{jk} for all $k = 1, \dots, n$, then we use the following compromise drift function decomposition:

$$P_{ij}(w, v) = P_{loc}(w, v) \prod_{k=1}^n P_{char}(w, v; c_{ik}, c_{jk}). \quad (10)$$

We use a characteristic function as our localisation function in this paper,

$$P_{loc}(w, v) = \mathbb{1}_{\{|w-v| < r\}},$$

for some $r > 0$. We use only P_{loc} for the leader species that is $P_{iL}(w, v) = P_{LL}(w, v) = P_{loc}(w, v)$.

We now discuss the nature of our leader species, in terms of political parties. Let there be \mathcal{P} parties, $\Pi_1, \dots, \Pi_{\mathcal{P}}$, in our political system, and $a = p_0 < p_1 < \dots < p_{\mathcal{P}-1} < p_{\mathcal{P}} = b$ be a partition of $\mathcal{I} = [a, b]$. Then we distribute the leader species population in the intervals $\mathcal{I}_\zeta = [p_{\zeta-1}, p_\zeta]$, which we call “influence intervals”. We run an “election” at time t , on the microscopic level, by assuming that all individuals vote for the party who controls the influence interval they are contained within, and on the mesoscopic and macroscopic levels, by computing “influence masses”,

$$M_{\zeta,i}(t) = \int_{\mathcal{I}_\zeta} f_\zeta(w, t) dw$$

for every $\zeta = 1, \dots, \mathcal{P}$, a graphical representation of which is presented in Figure 1.

The voter intention polling data we use is organised into tables by characteristics, with party on the vertical axis and the groups (e.g. leave and remain for 2016 Brexit Vote, 18-24 year-old etc. for age) on the horizontal axis. A table for an example characteristic k is displayed in Table 1 and the actual table for the 2016 Brexit vote is displayed in Table 2.

Table 1: Example of how the characteristic data is presented.

| Characteristic k | Group 1 | Group 2 | ... | Group m |
|---------------------|-----------------------|-----------------------|----------|-----------------------|
| Π_1 | $x_{1,1,k}$ | $x_{1,2,k}$ | ... | $x_{1,m,k}$ |
| Π_2 | $x_{2,1,k}$ | $x_{2,2,k}$ | ... | $x_{2,m,k}$ |
| \vdots | \vdots | \vdots | \ddots | \vdots |
| $\Pi_{\mathcal{P}}$ | $x_{\mathcal{P},1,k}$ | $x_{\mathcal{P},2,k}$ | ... | $x_{\mathcal{P},m,k}$ |

For each binary interaction (1)-(3) we can find the influence intervals \mathcal{I}_ζ to which w, v belong to, assume that $w \in \mathcal{I}_\mu$ and $v \in \mathcal{I}_\nu$, for some $\mu, \nu = 1, \dots, \mathcal{P}$. We look up the relevant values in the tables to get the values of $x_{\mu,i,k}$ and $x_{\nu,j,k}$. We want our characteristic comparison function P_{char} to represent the idea of net movement towards more

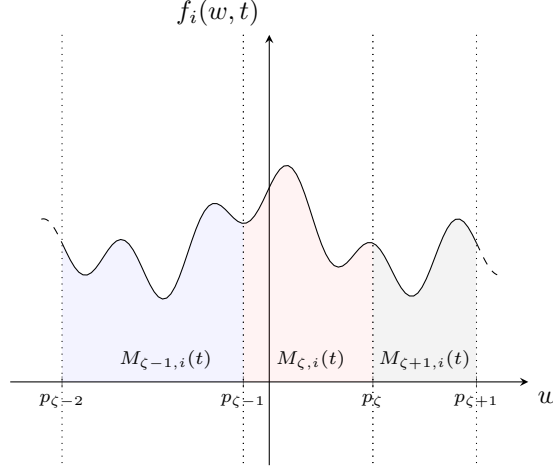


Figure 1: Schematic representation of influence masses for a given density function, $f_i(w, t)$, at some time t .

Table 2: The voter intention data for 2016 Brexit vote from [36].

| Brexit Vote | Leave | Remain |
|--------------|--------|--------|
| Labour | 0.1313 | 0.4081 |
| Green | 0.0202 | 0.0612 |
| LibDems | 0.0404 | 0.2551 |
| Other | 0.0303 | 0.0306 |
| Conservative | 0.6969 | 0.1938 |
| Brexit | 0.0606 | 0 |

popular opinions, popular in the sense of voter intentions from the pre-election polls. That is individuals with popular opinions move slower towards the unpopular opinions than the individuals with unpopular opinions move towards the more popular opinion. To this end, we let our characteristic comparison function depend on $x_{\mu,i,k} - x_{\nu,j,k}$, which gives a large, positive value when the party, in whose influence the individual resides, is more popular (i.e. $x_{\mu,i,k}$ dominates) and a large negative value when the party, in whose influence the other individual resides, is more popular (i.e. $x_{\nu,j,k}$ dominates). When both parties are equally popular, this value will be zero and therefore will not have a bias towards one party or the other. We are therefore using $x_{\mu,i,k} - x_{\nu,j,k}$ since it is representative of the popularity of the individual's opinion.

The value of $x_{\mu,i,k} - x_{\nu,j,k}$ is in $[-2, 2]$, but as previously stated our P -function's value must be in $[0, 1]$, hence we project the value of $x_{\mu,i,k} - x_{\nu,j,k}$ into $[0, 1]$. To achieve this we use the linear function,

$$H(x) = \frac{1}{4}(2 + x).$$

Since we are using $x_{\mu,i,k} - x_{\nu,j,k}$ as a measure of the popularity of the individual's opinion w , we want P_{char} to be decreasing with respect to its argument $x_{\mu,i,k} - x_{\nu,j,k}$, to model the decrease in movement from, and the increase in movement to, popular opinions. To this end we use a second function,

$$G(x) = \left(1 - x^a\right)^b,$$

for some $a, b > 0$. Note that G has the properties: (a) $G(0) = 1$ and $G(1) = 0$; and (b) this function is decreasing in $[0, 1]$, for any value of $a, b > 0$. These properties, in combination, lead to the value of G being close to zero when $x_{\mu,i,k}$ is dominant (or the individual holds a more popular opinion), and represents a small movement towards the other more unpopular opinion and vice versa. Finally, we compose these functions and thus we use the function,

$$F(x) = G(H(x)) = \left(1 - \left(\frac{1}{4}(2 + x)\right)^a\right)^b,$$

with $a, b > 0$. Thus our P_{char} function is,

$$P_{char}(w, v; c_{i,k}, c_{j,k}) = F(x_{\mu,i,k} - x_{\nu,j,k}),$$

when viewed on the microscopic level. In general, we can write this function as the sum of characteristic functions which sort v, w into their respective influence intervals as follows;

$$P_{char}(w, v; c_{i,k}, c_{j,k}) = \sum_{\mu, \nu=1}^{\mathcal{P}} \mathbb{1}_{w \in [p_{\mu-1}, p_{\mu}]} \mathbb{1}_{v \in [p_{\nu-1}, p_{\nu}]} F(x_{\mu,i,k} - x_{\nu,j,k}) \quad (11)$$

where $\mathbb{1}_{x \in A}$ denotes the characteristic function for the set A .

3 Numerical experiments

We separate the analysis of the methods into two types, the Monte Carlo simulation of the Boltzmann-type system (4)–(5), and the Finite Element Method for the Fokker-Plank system (6)–(7).

Throughout this section, we have chosen η_i all independent and identically distributed, in such a way that they can only attain two values, ± 0.01 , with equal probability, to ensure that the interactions remain in \mathcal{I} . Unless otherwise stated, we use the values of $a = 2$ and $b = 8$ in the P_{char} function in Section 2.2.3 since these have offered the best results, $\alpha_i = \beta_i = 1$ for all $i = 1, 2, \dots, N, L$. We construct our initial densities from available polling data and from population values for the specific region we wish to run simulations in. We do this by first finding the influence mass of the Party Π_{ζ} for a given species i , multiplying the relevant columns of our data tables, which, for a species whose members have characteristic k and are in Group ξ , is the ξ -th column from Table 1 and the relevant column from every characteristics table, element-wise. We will call the elements of the resulting column $X_{\zeta,i}$. Recall that by definition

$$M_i = \sum_{\zeta} M_{\zeta,i},$$

where we recall that $M_{\zeta,i}$ are the influence masses from Section 2.2.3. Given some distribution of mass between the follower species M_i , we can compute the initial influence masses for the parties in

$$M_{\zeta,i} = \frac{v_{\zeta} X_{\zeta,i}}{\sum_{\zeta} v_{\zeta} X_{\zeta,i}} M_i,$$

with v_{ζ} being the percentage of votes allocated to Π_{ζ} in some test election, either through the recorded results of some previous election or some pre-election polling data. We define the initial densities of the species as a weighted sum of characteristic functions in the following way

$$f_i(w, 0) = \sum_{\zeta} M_{\zeta,i} \mathbb{1}_{w \in \mathcal{I}_{\zeta}},$$

for all $\zeta = 1, 2, \dots, N$.

The initial density of the leader species is chosen as a weighted sum of delta distributions,

$$f_L(w, 0) = \sum_{k=1}^6 p_k \delta(w - L_k), \quad (12)$$

with p_k representing the percentage of the vote that each party attained in this, or another, test election, and L_k representing the party locations. In all the simulations presented here, we set the mass of the leader species to have 5% the total mass of the follower species, in a similar way to *Düring et al* [2]. For purposes of the discretisation in the Fokker-Plank FEM simulation, we approximated the delta functions in $f_L(w, 0)$ with the scaled mollifier functions

$$\phi(x) = \begin{cases} \exp\left(-\frac{(x/\varepsilon)^2}{1 - (x/\varepsilon)^2}\right), & \text{for } x \in [-1, 1], \\ 0, & \text{otherwise,} \end{cases}$$

with scale factor $\varepsilon \ll 1$.

3.1 Monte-Carlo simulations of the Boltzmann-type system

We perform a series of kinetic Monte Carlo simulations for the Boltzmann-type system (4)–(5). In these simulations, called Direct Simulation Monte Carlo (DSMC) or Bird's scheme, we choose individuals from the population non-exclusively, and perform the interaction rules (3), (2) and (1), depending on the species those individuals fell into. We run all simulations with $M = 10500$ total individuals, with 500 individuals in the Leader species and the remaining 10000 individuals distributed into follower species. We distribute these individuals using data from previous elections and population data.

Each time step in our simulation corresponds to M interactions and we run the simulation for 1000 time steps averaging over the last 50 time steps. We average the results of eight realisations to give a good approximation of the long-term density. We discretise the spatial domain $\mathcal{I} = [-1, 1]$ in 50 sub-intervals.

3.2 Finite Element discretisation of the Fokker-Planck system

Throughout this section, we use (f, g) to indicate the L^2 inner product of the function f and g , that is, $(f, g) = \int_{\mathcal{I}} f(w) \cdot g(w) dw$.

We discretise our time axis with $A = 300$ nodal points, and space axis with $B = 100$ nodal points. We write the nodal values of our functions $f_i(w_k, t^n) = f_{i,k}^n$ and the vector of every spatial nodal value, at time $t^n > 0$, as \mathbf{f}_i^n . We then use a semi-implicit Euler method to discretise our coupled system of non-linear, non-local PDEs in time and a finite element method (FEM) for discretisation in space, with quadratic, Lagrangian basis functions $b_k \in C_c(\mathcal{I}; \mathbb{R})$ for $k = 1, 2, \dots, B$, which leads to the following linear system:

$$\sum_{i=1}^N \left((\mathbf{M} + \mathbf{U}_i - \mathbf{C}_{Li}^n) \mathbf{f}_i^{n+1} + \sum_{j=1}^N (\mathbf{V}_{ij} - \mathbf{T}_{ij}^n) \mathbf{f}_j^{n+1} \right) = \sum_{i=1}^N \mathbf{M} \mathbf{f}_i^n, \quad (13)$$

where $\mathbf{M} = (b_k, b_l)_{(k,l)}$ is the mass matrix, \mathbf{C}_{Li}^n and \mathbf{T}_{ij}^n are the discretised transport matrices, given by

$$\mathbf{C}_{Li}^n = \Delta t \left(\left[\frac{1}{2\tau_{iL}} K_{L,iL}^n + \sum_{j=1}^N \frac{1}{2\tau_{ij}} K_{j,ij}^n \right] b_k, b'_l \right)_{(k,l)}, \quad \mathbf{T}_{ij}^n = \Delta t \left(\frac{\alpha_{ij}}{2\tau_{ij}} K_{i,ij}^n b_k, b'_l \right)_{(k,l)},$$

with $K_{L,iL}^n(w)$ and $K_{m,ij}^n(w)$ as the midpoint rule approximation of the convolution operators, $\mathcal{K}_{m,ij}(w, t^n)$ evaluated at the previous time step, in the transport terms of the Fokker-Planck equations and \mathbf{U}_i and \mathbf{V}_{ij} are the discretised Laplacian term matrices,

$$\mathbf{U}_i = \Delta t \left(\left[\frac{\lambda_i M_L}{4\tau_{iL}} + \sum_{j=1}^N \frac{\lambda_i M_j}{4\tau_{ij}} \right] D^2 b'_k, b'_l \right)_{(k,l)}, \quad \mathbf{V}_{ij} = \Delta t \left(\frac{\lambda_i M_j \beta_{ij}^2}{4\tau_{ij} \alpha_{ij}} D^2 b'_k, b'_l \right)_{(k,l)}.$$

3.3 Application to the 2019 General Election in the UK

We now consider an application of the above model on the United Kingdom General Election in 2019. The electoral areas of the United Kingdom are various local areas called constituencies, each of which elect a single member of parliament (MP) to the House of Commons by a first-past-the-post system, that is the candidate with the most votes in the constituency is declared the winner and becomes the MP. The political party with the largest number of seats in the House of Commons is, by convention, called the winning party and the Leader of the winning party is usually chosen to be the Prime Minister. We want to run simulations of this model on an individual constituency level, however the freely available voter intention data from YouGov we use [36] is presented on a national level for England, Scotland and Wales. We therefore approximate local data using the methods presented in this section. We choose to run simulations in various regions, 1) the Office of National Statistics' (ONS) regions of England, which are: the East of England, East Midlands, London, North East, North West, South East, South West, West Midlands, and Yorkshire and the Humber; and 2) the county of Nottinghamshire containing the constituencies Ashfield, Bassetlaw, Broxtowe, Gedling, Mansfield, Newark, Nottingham East, North and South, Rushcliffe, and Sherwood. We use the ONS' regions since they allow for results representative of the overall result in these large areas and the county of Nottinghamshire because Ashfield, Bassetlaw and Mansfield are part of the so-called "Red Wall", a group of constituencies in the North and Midlands of the UK who traditionally elected Labour MPs, yet elected many Conservative MPs in the 2019 General Election, in what was considered to be a decisive factor in the overall outcome of the election [37, 38]. We use the 2011 Census of the UK [39] to get the total population and the demographic population for the above areas.

We have made the assumption that supporters of the UK independence party (UKIP) from 2015 and 2017 General Elections [40, 41] are aligned closer to the political views of the Brexit Party in the 2019 General Election. This is because the manifesto of the Brexit Party is very similar to that of UKIP in those years. In the voter intention data from YouGov, the parties of Plaid Cymru and the Scottish National Party (SNP) are included but since they do not run a candidate in any constituency in the East of England, nor in Nottinghamshire, we have omitted them from the simulations.

For these simulations, we focus on two characteristics, age, split into 18–24 year-old, 25–49 year-old, 50–64 year-old and 65+ year-old, and self-reported 2016 Brexit Vote, split into Leave and Remain. Thus our eight follower species will be 18-24 year-olds, who voted leave, 25-49 year-olds who voted leave etc. We again reiterate that since we are modelling this process over a number of years, there will be some demographic shift as individuals at the upper end of an age range will shift to a different species, however it may still be reasonable not to take this into account here, if these individuals do not exhibit a large change in political view over this period. If required, demographic changes (migration to other species, birth/death) could be incorporated by adding gain and loss terms similarly as in [2].

We choose our party locations, that is the values of L_k in (12), to be: Labour at -0.7 , Green at -0.4 , Liberal Democrats (LibDems) at -0.1 , Other at 0.1 , Conservative at 0.4 , and the Brexit Party at 0.7 . This leads us to an intuitive way to choose the influence intervals, \mathcal{I}_k , by choosing the midpoint between the party locations as the partitions, as well as -1 and 1 for the most extreme opinions: Labour has influence in the region $[-1, -0.55)$, Green in $[-0.55, -0.25)$, LibDems in $[-0.25, 0)$, Other in $[0, 0.25)$, Conservatives in $[0.25, 0.55)$, and the Brexit Party in $[0.55, 1]$. In the FEM of the Fokker-Planck system, we will use the scale factors for our mollifier functions as $\varepsilon = 0.08$, so that the peaks that are associated with the parties remains entirely within that parties influence interval.

3.3.1 Case study: East of England

In this subsection, we look specifically at the East of England region. However, every step we use in this section, can be applied similarly to all areas described above.

To process the data, we take the results for the 2017 General Election [41] and compute the total vote tallies for the East of England. This gives us the popular vote percentages in the region, the result of which is displayed in Table 3.

Table 3: Popular vote percentage for the East of England for the 2017 General Election.

| | proportion of vote |
|--------------|--------------------|
| Labour | 0.3271 |
| Green | 0.0189 |
| LibDems | 0.0788 |
| Other | 0.0034 |
| Conservative | 0.5464 |
| UKIP | 0.0251 |

We then use the demographic voter intention data from YouGov [36], and multiply this by the previous election result of the region and normalise it as seen in Table 4. We compute the masses of the species densities M_i to reflect

Table 4: YouGov data (left) and regionalised data (right) for the Brexit vote for the East of England.

| | YouGov national | | East regionalised | |
|--------------|-----------------|--------|-------------------|--------|
| | Leave | Remain | Leave | Remain |
| Labour | 0.1313 | 0.4081 | 0.1001 | 0.5119 |
| Green | 0.0202 | 0.0612 | 0.0009 | 0.0044 |
| LibDems | 0.0404 | 0.2551 | 0.0074 | 0.0771 |
| Other | 0.0303 | 0.0306 | 0.0002 | 0.0004 |
| Conservative | 0.6969 | 0.1938 | 0.8877 | 0.4061 |
| Brexit | 0.0606 | 0 | 0.0035 | 0 |

the population size in the 2011 Census which is already split by region and age, then divide them based on the result of 2016 European Union Referendum in the East of England region, the data for which is freely available from the

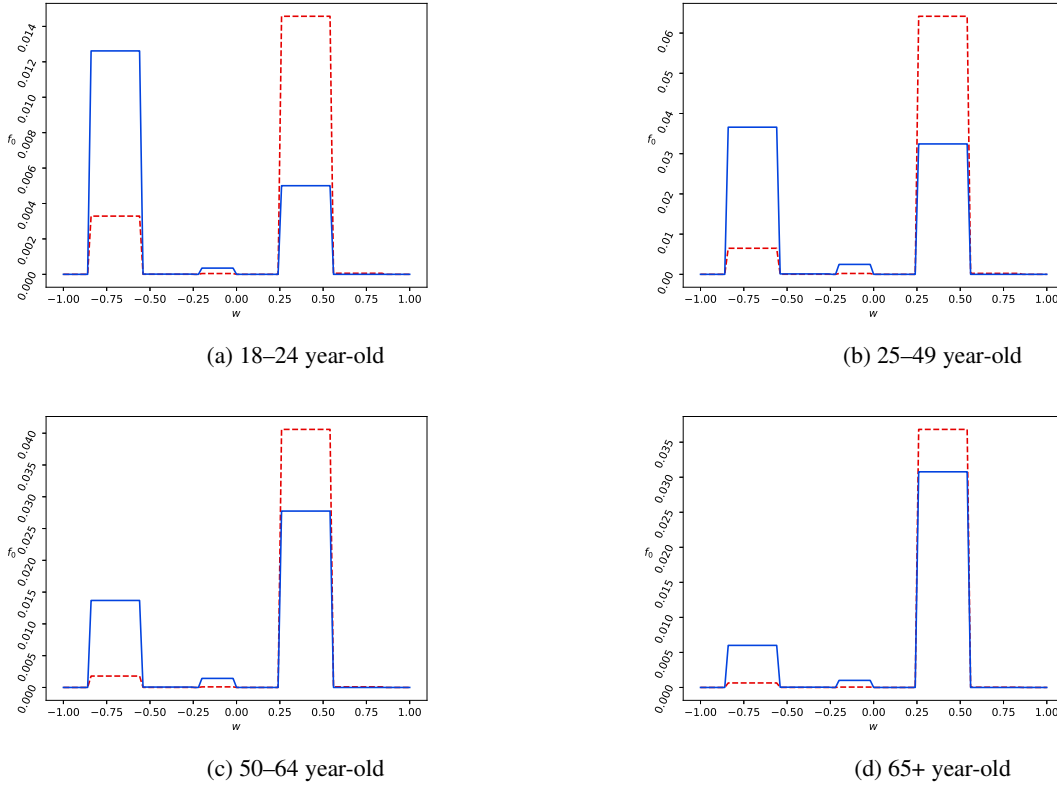


Figure 2: Initial follower densities for the eight follower species, from 2015 General Election Results. Each graph shows the initial density for the leave (solid blue) and the remain (dashed red) characteristic of an age group.

House of Commons Library [42]. For example, to compute the approximate number of 18-24 year-olds who voted to leave, we find the percentage of the electorate of the East of England that is 18-24 years-old,

$$\frac{\text{\# of 18-24 year-olds}}{\text{\# of eligible voters in EOE}} = \frac{490,197}{4,617,534} = 0.1061,$$

then we compute the percentage of the electorate that voted to leave in the 2016 Brexit vote in the East of England,

$$\frac{\text{\# of voters that voted leave}}{\text{\# of valid votes cast}} = \frac{1880367}{3328983} = 0.5648.$$

We approximate the mass of the population that is 18-24 years-old and voted leave by taking the product of the above values, that is $0.1061 \times 0.5648 = 0.0599$. We repeat this for all species, and we obtain the values in Table 5. We use the method from Section 3, using the voter intention data from YouGov and the 2015 General Election Results to generate our initial conditions for the follower species, as described in Section 3. A graphical representation of these initial functions $f_i(w, 0)$ is given in Figure 2 where we have displayed the eight initial follower densities by age in subgraphs and by self-reported Brexit vote by colour.

For the leader species initial condition, we use directly the results from the 2015 General Election and the graph of the resultant initial function, $f_i(w, 0)$ in the East of England Region used in the Fokker-Plank simulation is presented in Figure 3.

3.4 Discussion of results

We present here the results from our numerical experiments of the model presented thus far. First we look at the how the DSMC and FEM methods compare when applied to the East of England, as described in Section 3.3.3, and lastly in this section, we investigate the effect of the P_{char} function on the model by comparing the choice from Section 2.2.3 to a model analogous with that of *Düring et al.*. We use the FEniCS Project [43, 44], to simulate the Fokker-Plank system, with graphs drawn using Matplotlib [45] and the DSMC experiments are run using Python 3. The parameters

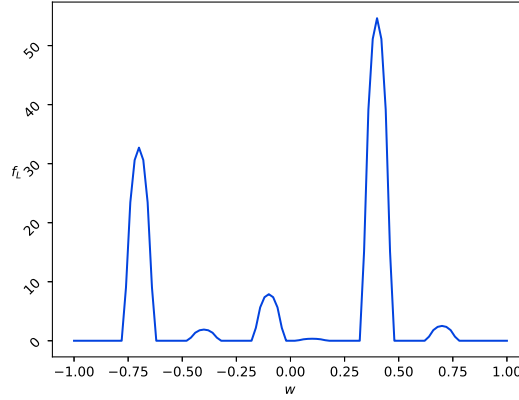


Figure 3: The mollified, initial Leader species density for the East of England Region.

 Table 5: The masses M_i of the follower species i in the East of England.

| | Leave | Remain |
|-------|--------|--------|
| 18-24 | 0.0599 | 0.0461 |
| 25-49 | 0.2374 | 0.1829 |
| 50-64 | 0.1421 | 0.1094 |
| 65+ | 0.1252 | 0.0965 |

used for both models are similar to those used in [2, 46]. In the Fokker-Planck system larger (smaller) values of λ_i do smooth out profiles more (less), but do not affect the positions of local maxima. The choices of γ_i and η affect the results of the DSMC simulation in a similar way (since $\lambda_i = \sigma_i^2/\gamma_i$ in the limit process in Section 2.2.2). The choice of r determines the range of interactions, and hence smaller values of r lead to a larger number of local maxima, compare the results in [24].

3.4.1 Numerical Results: comparing Boltzmann-type to Fokker-Planck

First, we run the DSMC experiments for the Boltzmann-type system and for the Fokker-Planck system with the same – or equivalent – parameters. We do this to illustrate that the numerical results from the Boltzmann-type system and the Fokker-Planck system are good approximations of each other. We choose our parameters in the simulations as: $r = 0.5$, $\gamma_i = \gamma = 0.3$, $\eta = \pm 0.01$ each with probability $1/2$, $\gamma_L = 0$ and $D_L(w) \equiv 0$ in the DSMC for the Boltzmann-type system; and $r = 0.5$, $\tau_{iL} = 0.05$, $\tau_{ij} = 2.5$, $\lambda_i = 0.033$, $\Delta s = 0.1$ in the FEM for the Fokker-Planck system.

The results of this comparative simulation are presented in Figure 4. As we can observe, the results of the DSMC (the blue bars) and the FEM (the red dashed lines) coincide well. This suggests that using the FEM of the Fokker-Planck system is reasonable. We therefore present only the numerical experiments of the Fokker-Planck system in the remainder of this section.

3.4.2 Numerical Results: investigating the effect of data-driven P_{char}

In our second numerical experiment, we illustrate the effect of our choice of P_{char} which is chosen from demographic voter intention data (see Section 2.2.3) when compared to a model with $P_{char} \equiv 1$, similar to that presented by Düring *et al.* [2]. To this end, we run various Fokker-Planck simulations with uniform initial density, with species' masses (see Table 5 in Section 3), looking at the long-time densities for $\tau_{iL} = 0.05$, $\tau_{ij} = 2.5$, $\lambda = 0.033$, $\Delta s = 0.1$ and $r = 0.85$. We run the model for 300 time steps to ensure the resulting distribution is the limiting distribution. We expect the formation of at least two peaks from our model, one within the influence of Labour ($[-1, -0.55]$) and one within the Conservative Influence ($[0.25, 0.55]$). We also expect that, in our model, the species have distinct, yet similar, densities whereas in the model with $P_{char} \equiv 1$, we expect the densities to be identical. This is confirmed in our numerical simulations in Figure 5.

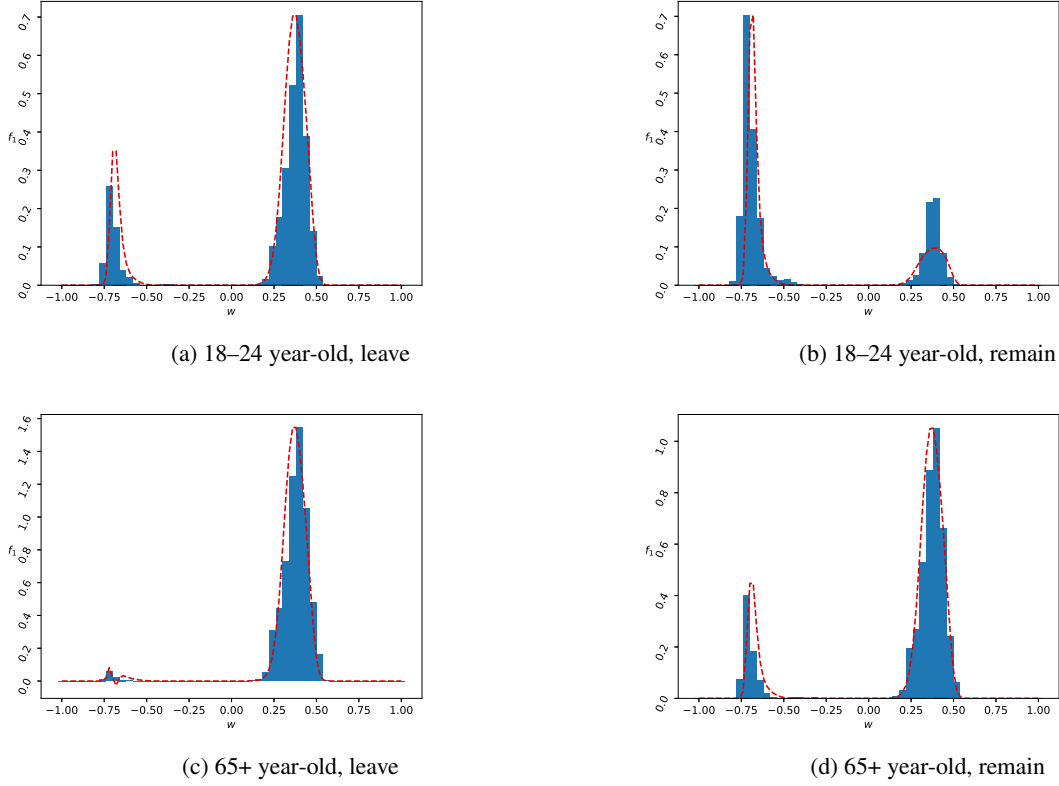


Figure 4: Numerical simulation of 2019 UK General Election in East of England: histogram representing the distribution of the individuals in the DSMC, with the results from a similar Fokker-Planck FEM solution superimposed as red dashed lines.

In this simulation, we observe that a small peak appears in the influence of Labour in the case where our choice of P_{char} is used, and the difference between the species' densities is very similar but there is a small, but distinct difference between Leave and Remain distributions, displayed as red and blue.

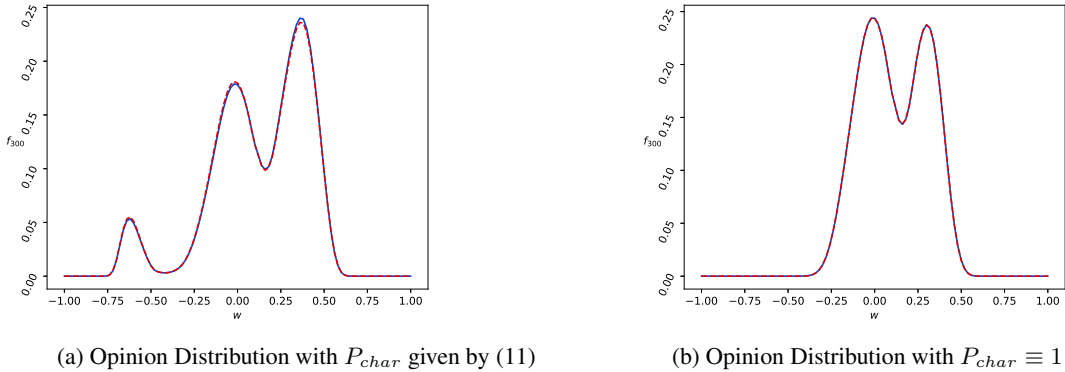


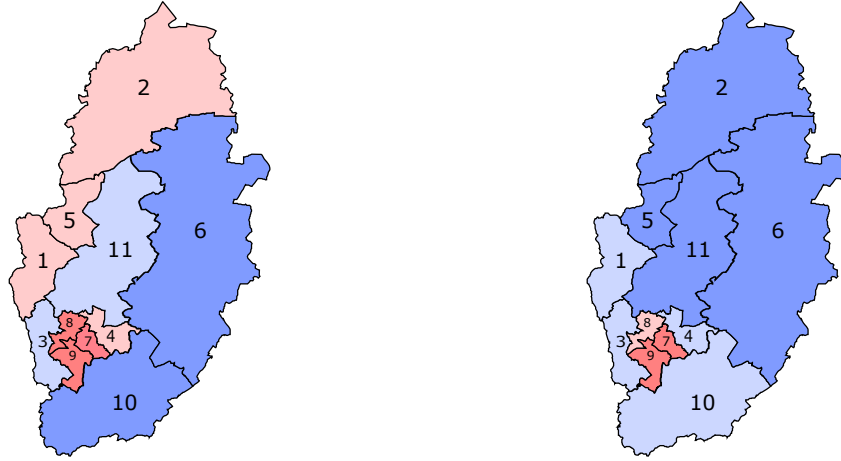
Figure 5: A comparison between the 18-24 year-old species in the East of England. (5a) are the results of the simulation with our choice of P_{char} , and (5b) are the results from the simulation with $P_{char} \equiv 1$.

From these simulations it is clear that the advection terms, through the \mathcal{K} operators, are dominant in this type of model and so we see a net movement towards the large parties, from the compromise function P_{char} and towards the centre of the domain, from the localisation function P_{loc} . We believe that advection towards the centre of the two popular peaks at Labour (-0.7) and the Conservatives (0.4), that arises from the localisation function, leads to the

success of the central parties. This then leads to the eclipse of the smaller parties, yielding only three peaks. This effect may be removed with the use of a smaller interaction radius r and rearranging of the parties.

3.5 Numerical results for 2019 UK general election’s “Red Wall”

Finally, we apply this model to a real-world situation, to show a proof of concept that models of this type can be used with sufficient data. As discussed in Section 3.3.3, we run Fokker-Planck simulations for the county of Nottinghamshire using the parameter values from the simulations in the East of England simulations above namely: $\tau_{iL} = 0.05$, $\tau_{ij} = 2.5$, $\lambda = 0.033$, $\Delta s = 0.1$ and $r = 0.85$, where we use the same method presented in Section 3.3.3(3.3.1), to generate our initial masses, namely by taking the product of the relevant columns in our data tables (an example of which is Table 1) element-by-element for both age and self-reported 2016 Brexit vote, then dividing this mass evenly between the populations obtained from the 2011 Census, and to regionalise the voter intention data, by taking the columns of the national voter intention data and the results from the 2017 General Election and multiplying element-by-element. We also generate the initial densities in the same way as that section, that is we use sums of characteristic functions. The Nottinghamshire constituencies in Figures 6 and 7 are labeled by numbers: (1) Ashfield, (2) Bassetlaw, (3) Broxtowe, (4) Gedling, (5) Mansfield, (6) Newark, (7), (8) and (9) Nottingham East, North and South respectively, (10) Rushcliffe and (11) Sherwood. Colour-shading indicates the outcome of the 2015 and 2019 General Election in each constituency.



(a) Result of the 2015 General Election in Nottinghamshire (b) Result of the 2019 General Election in Nottinghamshire

Figure 6: Results of the General Elections in 2015 (6a) and 2019 (6b). The colours represent the party to which the elected MP belongs with blue for Conservative and red for Labour. The light colours are where the total popular vote for the elected MP was below 50% and the dark colours for greater than 50%.

We see the shift of the northernmost three constituencies, Ashfield, Bassetlaw and Mansfield, from Labour, as a part of the “Red Wall”, to Conservative in Figure 6 in the 2019 General Elections. The so-called “fall of the Red Wall” was seen as a decisive factor in the 2019 victory for the Conservatives [37, 38] since many Labour MPs were elected in these constituencies, forming the backbone of the Labour party throughout most of the 20th century.

This is a good example for testing our new model since, from the recorded results of the 2015 and 2019 General elections [41, 47], there are a few constituencies which elected a Labour MP in 2015 and a Conservative MP in 2019 (Ashfield, Bassetlaw, Mansfield and Gedling), a few constituencies that elected a Conservative MP in both elections (Broxtowe, Newark, Rushcliffe and Sherwood), and some that elected a Labour MP in both elections (Nottingham East, North and South).

As can be seen in the numerical simulation results in Figure 7a, our model correctly assigns a “winner”, the party with the highest influence mass (aggregated over all eight demographic species), in each constituency. The results from the kinetic model predict a larger winning majority than the actual results in some constituencies, however, these quantitative differences are less relevant in these first-past-the-post elections, demonstrating the utility of this model. We believe that further refinement of this model along with more current data may alleviate this issue, which we leave to future research.

We now look more in depth at the constituency of Bassetlaw, for which the initial eight follower densities (generated from the 2015 General Election Results) are presented in Figure 8, and the predicted results from our model are

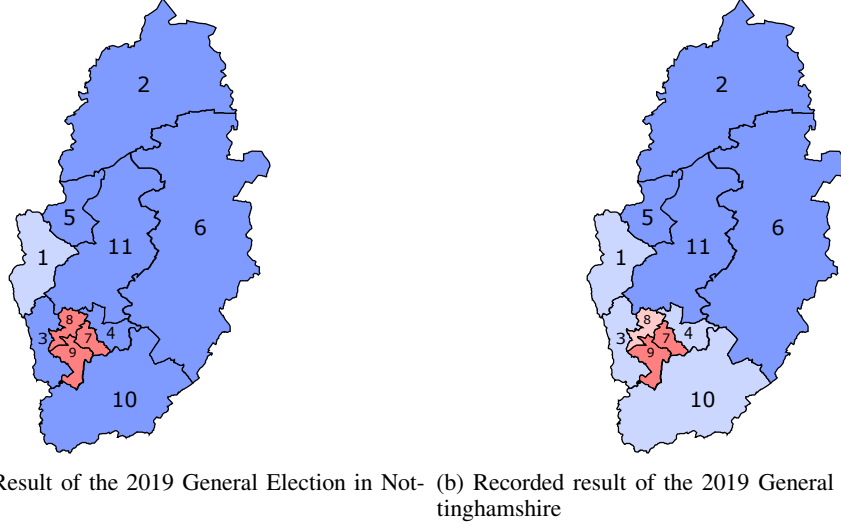


Figure 7: Numerical simulation results predicted by our proposed kinetic model, compared to the 2019 General Election result in Nottinghamshire. The colours represent the party to which the elected MP belongs with blue for Conservative and red for Labour. The light colours are where the total popular vote for the elected MP was below 50% and the dark colours for greater than 50%.

presented in Figure 9. As we can see, the distinction between the different demographics is quite clear, the younger demographics – those with ages 18-24 years-old and 25-49 years-old – show the largest difference in initial to predicted opinion, whereas the two older demographics show little change from the initial density. Since the demographics with the largest total masses are the 25-49 year-olds (both leave and remain), this shift mass from the Labour influence interval to the Conservative influence interval may have been the cause for the entire constituency to vote for a Conservative MP in the 2019 General Election.

We additionally note that the older demographics masses are either split somewhat evenly over the Labour and Conservative influence intervals (in the case of Remain voters) or the majority of the mass in the Conservative influence interval (in the case of Leave voters) as compared to the higher mass in the Labour influence interval in general for the younger demographics. Figure 9 shows a clear correlation between age and Conservative voting and self-reported Brexit vote and Conservative voting predicted in our model – a correlation that has been observed since the election [48].

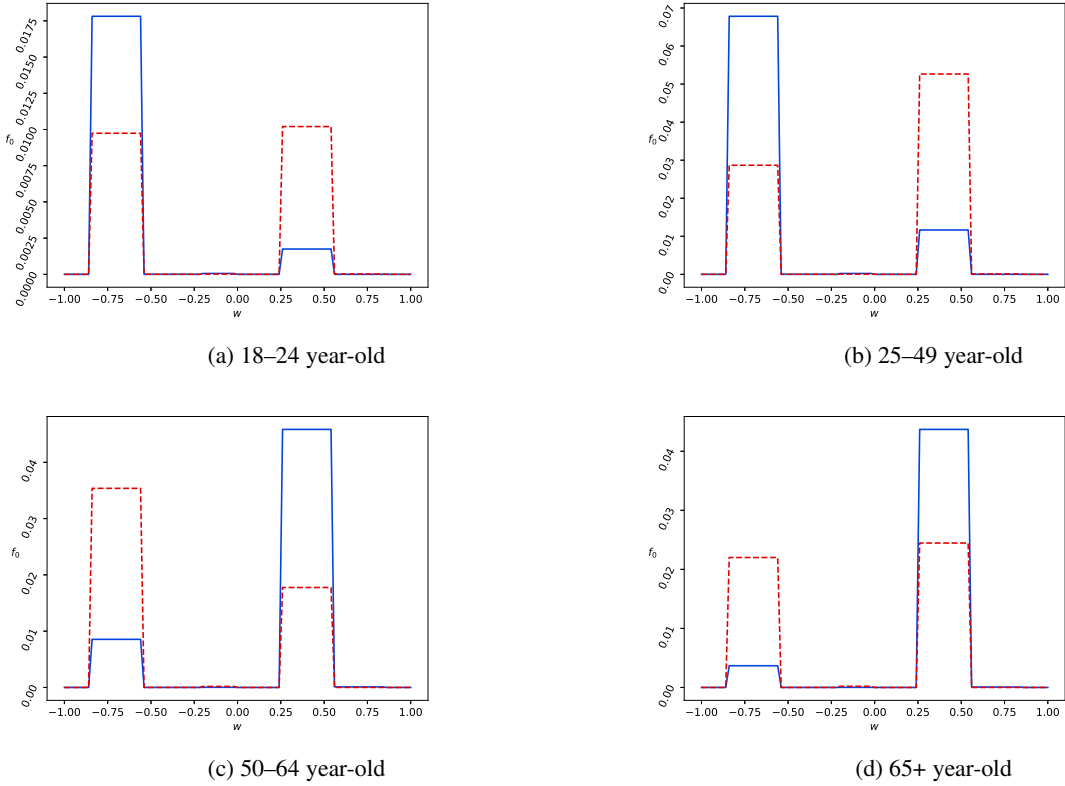


Figure 8: Initial densities for the eight follower species in the constituency of Bassetlaw from the 2015 General Election Results. Each graph shows the initial density for the leave (solid blue) and the remain (dashed red) characteristic of an age group.

4 Conclusion

We have proposed a kinetic model for opinion formation which takes into account the effect of voter demographics and other socio-economic factors on the opinion formation dynamics. Using methods of kinetic theory, from microscopic opinion consensus dynamics, we have derived a mesoscopic Boltzmann-type system and, in a quasi-invariant limit, a macroscopic Fokker-Planck system. In this model, the followers' compromise process is biased towards their preferred opinions, influenced by their demographic factors like age and their behaviour in previous election or referenda, as suggested by the results of real-world voter polls and post-election analyses. The use of voter intention data in the compromise function yields interpretable model parameter functions, unlike in alternative approaches, as for example a process of calibrating mathematical model parameters to obtain (or be close to) a desired target distribution. Application of this model to the 2019 UK general election, using voter intention data and previous election data illustrates the potential of such models and also highlights the importance of quality data. For Nottinghamshire, part of the so-called "Red Wall", where there was a strong shift in traditional Labour strongholds to the Conservative party, the model has been able to correctly assign all parliament seat winners. This election outcome is explained among other factors by the strong support for "Leave" in the 2016 Brexit referendum vote in these constituencies and the positioning of the two large parties in the EU membership question – an effect which our model is able to capture and include in the opinion formation dynamics.

Data Access

YouGov voter intention data is available at [36]. 2011 Census data is available at [39]. Results of the 2015, 2017, 2019 UK general elections are available at [40, 41, 47]. Results of the 2016 EU referendum are available at [42]. The code used to run the simulations in Section 3 can be found at: <https://github.com/the-ollie-wright/kofm-pep>.

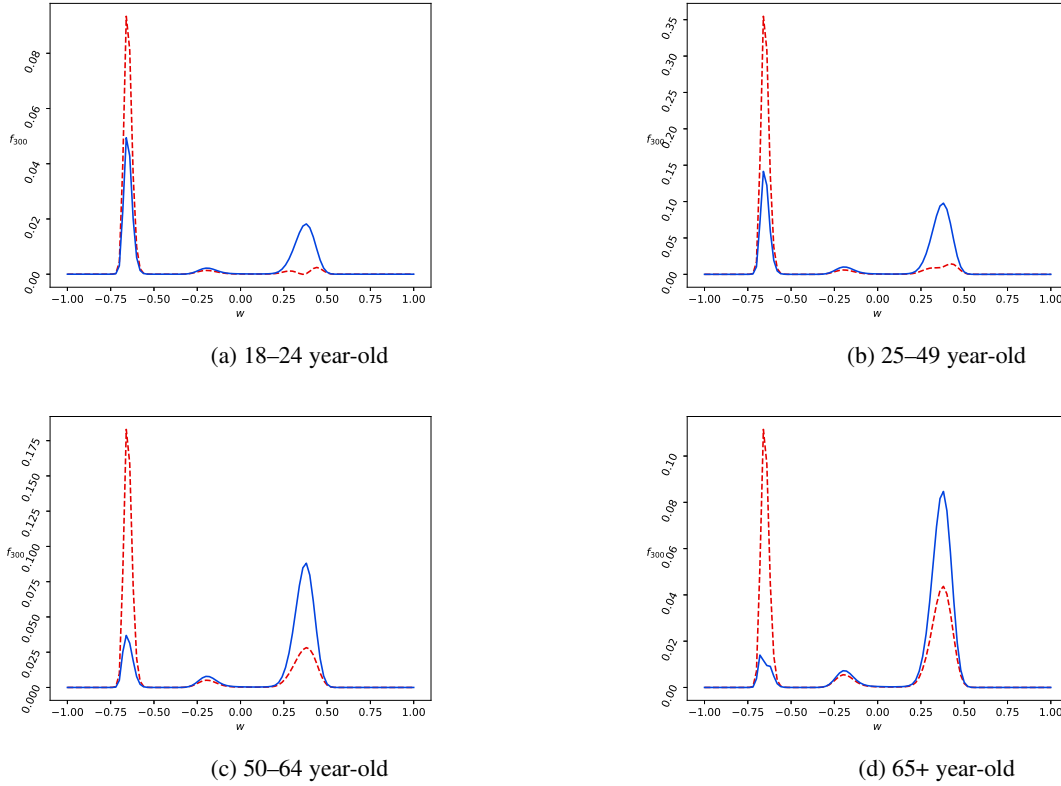


Figure 9: Predicted Results from the model for the constituency of Bassetlaw. Each graph shows the predicted density for the leave (solid blue) and the remain (dashed red) characteristic of an age group.

References

- [1] G. Toscani. “Kinetic models of opinion formation”. *Comm. Math. Sci.* 4 (3) (2006), pp. 481–496. DOI: 10.4310/CMS.2006.v4.n3.a1.
- [2] B. Düring, P. Markowich, J. Pietschmann, and M. Wolfram. “Boltzmann and Fokker-Planck equations modelling opinion formation in the presence of strong leaders”. *Proc. R. Soc. A.* 465 (2009), pp. 3687–3708. DOI: 10.1098/rspa.2009.0239.
- [3] M. L. Bertotti and M. Delitala. “On a discrete generalized kinetic approach for modelling persuader’s influence in opinion formation processes”. *Math. comput. model.* 48 (7-8) (2008), pp. 1107–1121. DOI: 10.1016/j.mcm.2007.12.021.
- [4] S. Motsch and E. Tadmor. “Heterophilous dynamics enhances consensus”. *SIAM Rev.* 56 (4) (2014), pp. 577–621. DOI: 10.1137/120901866.
- [5] P.-E. Jabin and S. Motsch. “Clustering and asymptotic behavior in opinion formation”. *J. Differ. Equations* 257 (11) (2014), pp. 4165–4187. DOI: 10.1016/j.jde.2014.08.005.
- [6] B. Düring and M. Wolfram. “Opinion dynamics: inhomogeneous Boltzmann-type equations modelling opinion leadership and political segregation”. *Proc. R. Soc. A.* 471 (2015), p. 20150345. DOI: 10.1098/rspa.2015.0345.
- [7] G. Albi, L. Pareschi, and M. Zanella. “Opinion dynamics over complex networks: kinetic modeling and numerical methods”. *Kinet. Relat. Models* 10 (1) (2017), pp. 1–32. DOI: 10.3934/krm.2017001.
- [8] G. Toscani and L. Pareschi. *Interacting Multiagent Systems*. Oxford University Press, 2013.
- [9] G. Toscani, A. Tosin, and M. Zanella. “Opinion modeling on social media and marketing aspects”. *Phys. Rev. E* 98 (2) (2018), p. 022315. DOI: 10.1103/PhysRevE.98.022315.
- [10] G. Albi, L. Pareschi, and M. Zanella. “On the optimal control of opinion dynamics on evolving networks”. In: *IFIP Conference on System Modeling and Optimization*. Springer, 2015, pp. 58–67. DOI: 10.1007/978-3-319-55795-3_4.

- [11] G. Albi, L. Pareschi, G. Toscani, and M. Zanella. “Recent advances in opinion modeling: control and social influence”. In: *Active Particles, Volume 1. Modeling and Simulation in Science, Engineering and Technology*. Ed. by T. E. Bellomo N. Degond P. Birkhäuser, Cham, 2017. DOI: 10.1007/978-3-319-49996-3_2.
- [12] B. Düring, A. Jüngel, and L. Trussardi. “A kinetic equation for economic value estimation with irrationality and herding”. *Kinet. Relat. Models* 10 (1) (2017), pp. 239–261. DOI: 10.3934/krm.2017010.
- [13] N. Bellomo, F. Colasuonno, D. Knopoff, and J. Soler. “From a systems theory of sociology to modeling the onset and evolution of criminality”. *Netw. Heterog. Media* 10 (3) (2015), pp. 421–441. DOI: 10.3934/nhm.2015.10.421.
- [14] G. Furioli, A. Pulvirenti, E. Terraneo, and G. Toscani. “Fokker-Planck equations in the modeling of socio-economic phenomena”. *Math. Models Methods Appl. Sci.* 27 (1) (2017), pp. 115–158. DOI: 10.1142/S0218202517400048.
- [15] P. Degond, J.-G. Liu, S. Merino-Aceituno, and T. Tardiveau. “Continuum dynamics of the intention field under weakly cohesive social interaction”. *Math. Models Methods Appl. Sci.* 27 (1) (2017), pp. 159–182. DOI: 10.1142/S021820251740005X.
- [16] B. Chazelle, Q. Jiu, Q. Li, and C. Wang. “Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics”. *J. Differ. Equations* 263 (1) (2017), pp. 365–397. DOI: 10.1016/j.jde.2017.02.036.
- [17] J. Garnier, G. Papanicolaou, and T.-W. Yang. “Consensus convergence with stochastic effects”. *Vietnam J. Math.* 45 (1-2) (2017), pp. 51–75. DOI: 10.1007/s10013-016-0190-2.
- [18] E. Cristiani and A. Tosin. “Reducing complexity of multiagent systems with symmetry breaking: an application to opinion dynamics with polls”. *Multiscale Model. Simul.* 16 (1) (2018), pp. 528–549. DOI: 10.1137/17M113397X.
- [19] M. Burger, M. Di Francesco, S. Fagioli, and A. Stevens. “Sorting phenomena in a mathematical model for two mutually attracting/repelling species”. *SIAM J. Math. Anal.* 50 (3) (2018), pp. 3210–3250. DOI: 10.1137/17M1125716.
- [20] L. Boudin and F. Salvarani. “Opinion dynamics: kinetic modelling with mass media, application to the Scottish independence referendum”. *Physica A* 444 (2016), pp. 448–457. DOI: 10.1016/j.physa.2015.10.014.
- [21] L. Pedraza, J. P. Pinasco, and N. Saintier. “Measure-valued opinion dynamics”. *Math. Models Methods Appl. Sci.* 30 (2) (2020), pp. 225–260. DOI: 10.1142/S0218202520500062.
- [22] M. Pérez-Llanos, J. P. Pinasco, N. Saintier, and A. Silva. “Opinion formation models with heterogeneous persuasion and zealotry”. *SIAM J. Math. Anal.* 50 (5) (2018), pp. 4812–4837. DOI: 10.1137/17M1152784.
- [23] G. Albi, L. Pareschi, and M. Zanella. “Boltzmann games in heterogeneous consensus dynamics”. *J. Stat. Phys.* 175 (1) (2019), pp. 97–125. DOI: 10.1007/s10955-019-02246-y.
- [24] L. Pareschi, G. Toscani, A. Tosin, and M. Zanella. “Hydrodynamic models of preference formation in multi-agent societies”. *J. Nonlinear Sci.* 29 (6) (2019), pp. 2761–2796. DOI: 10.1007/s00332-019-09558-z.
- [25] M. Burger. “Network structured kinetic models of social interactions”. *Vietnam J. Math.* (2021). DOI: 10.1007/s10013-021-00505-8.
- [26] B. Düring, D. Matthes, and G. Toscani. “Kinetic equations modelling wealth redistribution: A comparison of approaches”. *Phys. Rev. E* 78 (5 Nov. 2008), p. 056103. DOI: 10.1103/PhysRevE.78.056103.
- [27] S. Galam. “Sociophysics: A review of Galam models”. *Int. J. Mod. Phys. C* 19 (03) (2008), pp. 409–440. DOI: 10.1142/S0129183108012297.
- [28] G. Tett. “Can we fix polling? Yes. No. Not sure...” *The Financial Times* (2015). Last accessed 1 July 2021. URL: <https://www.ft.com/content/1b015de6-8e34-11e5-8be4-3506bf20cc2b>.
- [29] D. K. Park, A. Gelman, and J. Bafumi. “Bayesian multilevel estimation with poststratification: State-level estimates from national polls”. *Political Anal.* 12 (4 2004), pp. 375–385. DOI: 10.1093/pan/mp024.
- [30] C. Cercignani, R. Illner, and M. Pulvirenti. *The mathematical theory of dilute gases*. Vol. 106. Springer Science & Business Media, 2013.
- [31] R. Hegselmann and U. Krause. “Opinion dynamics and bounded confidence models, analysis, and simulation”. *J. Artif. Soc. Soc. Simul.* 5 (3) (2002).
- [32] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. “How can extremism prevail? A study based on the relative agreement interaction model”. *J. Artif. Soc. Soc. Simul.* 5 (4) (2002).
- [33] W. Weidlich. *Sociodynamics: A systematic approach to mathematical modelling in the social sciences*. Courier Corporation, 2006.
- [34] B. E. Lauderdale, D. Bailey, J. Blumenau, and D. Rivers. “Model-based pre-election polling for national and sub-national outcomes in the US and UK”. *Int. J. Forecast.* 36 (2) (2020), pp. 399–413. DOI: 10.1016/j.ijforecast.2019.05.012.

- [35] C. Cercignani. *Rarefied Gas Dynamics*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2000.
- [36] YouGov. *Voting intention by class identification*. Last accessed 1 March 2021. 2019. URL: <https://yougov.co.uk/topics/politics/survey-results>.
- [37] C. Bickerton. “Labour’s lost working-class voters have gone for good”. *The Guardian* (2019). Last accessed 1 July 2021. URL: <https://www.theguardian.com/commentisfree/2019/dec/19/labour-working-class-voters-bre>
- [38] BBC News. *General election 2019: Conservatives take Wakefield from Labour*. Last accessed 1 July 2021. 2019. URL: <https://www.bbc.co.uk/news/election-2019-50772224>.
- [39] Office of National Statistics. *2011 Census, KS102EW Age structure, local authorities in England and Wales*. Last accessed 1 March 2021. 2012. URL: <http://www.nomisweb.co.uk/census/2011/ks102ew>.
- [40] House of Commons Library. *General Election 2015 Full Results [Microsoft Excel Spreadsheet]*. Last accessed 5 March 2021. 2015. URL: <https://commonslibrary.parliament.uk/research-briefings/cbp-7186/>.
- [41] House of Commons Library. *Data file: detailed results by constituency [Microsoft Excel Spreadsheet]*. Last accessed 5 March 2021. 2019. URL: <https://commonslibrary.parliament.uk/research-briefings/cbp-7979>.
- [42] The Electoral Commission. *Results and turnout at the EU referendum*. Last accessed 16 April 2021. 2019. URL: <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-e>
- [43] A. Logg, K.-A. Mardal, G. N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012. DOI: 10.1007/978-3-642-23099-8.
- [44] M. S. Alnæs et al. “The FEniCS Project Version 1.5”. *Arch. Num. Soft.* 3 (100) (2015). DOI: 10.11588/ans.2015.100.20553.
- [45] J. D. Hunter. “Matplotlib: A 2D graphics environment”. *Comput. Sci. Eng.* 9 (3) (2007), pp. 90–95. DOI: 10.5281/zenodo.2577644.
- [46] B. Düring, M. Torregrossa, and M. Wolfram. “Boltzmann and Fokker-Planck Equations Modelling the Elo Rating System with Learning Effects”. *J. Nonlinear Sci.* 29 (2019), pp. 1095–1128. DOI: 10.1007/s00332-018-9512-8.
- [47] House of Commons Library. *Data file: detailed results by constituency [Microsoft Excel Spreadsheet]*. Last accessed 9 June 2021. 2021. URL: <https://commonslibrary.parliament.uk/research-briefings/cbp-8749>.
- [48] YouGov. *How Britain voted in the 2019 general election*. Last accessed 12 October 2021. 2019. URL: <https://yougov.co.uk/topics/politics/articles-reports/2019/12/17/how-britain-voted-2019-general-el>